

EXPLORING THE MALAYSIAN ENGLISH NEWSPAPER CORPUS FOR LEXICOGRAPHIC EVIDENCE

Tan Siew Imm

National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616

Email: siewimm.tan@nie.edu.sg

This paper describes the analysis of the Malaysian English Newspaper Corpus for lexicographic evidence, and demonstrates the use of corpus data and information obtained from secondary sources in creating dictionary entries for 457 Malaysian English loanwords, compound blends, loan translations and lexical creations. Consisting primarily of nouns, these features include food names, terms for festivals and religious practices, names of social and recreational activities, and honorific titles and terms of respect. These features are produced and understood by the community, and are generally considered acceptable for use across a wide range of intranational domains. The dictionary entries created through the project described in this article will require revisions as more corpora, representing diverse registers of Malaysian English, become available. New lexical features will also need to be codified. The influences of languages other than Malay and Chinese should also be examined. Although core English words should be central to a Malaysian English dictionary, how some of these words have been adapted in the local sociolinguistic context will also have to be explored. These dictionary entries should be relevant to the educational and communicational needs of the society while maintaining an endonormative standard. It is argued that the codification of the vocabulary of Malaysian English can be an effective way to signal the legitimacy and the coming of age of this variety.

Keywords: Malaysian English; Malaysian English Newspaper Corpus; lexicography; endonormative standard

INTRODUCTION

This article focuses on key issues surrounding the compilation of dictionary entries for 457 localised words in Malaysian English (henceforth, ME). Comprising loanwords, compound blends, loan translations and lexical creations, these features reflect the influences of Malay and Chinese, and the creativity of ME users. They fulfil important sociocultural needs of the community and enhance ME users' capacity to assert various identities and to express courtesy, solidarity and friendship.

The lexicographic evidence used to compile these dictionary entries was drawn from the Malaysian English Newspaper Corpus (henceforth, MEN Corpus), a five-million-word corpus of newspaper texts constructed to serve as a tool to investigate contact-induced change in Malaysian English (for details, see Tan, 2013). The use of a newspaper corpus to study variation and change in the English language is not without precedent (see, for instance, Westin, 2002 and Levin, 2006). Although newspapers represent merely a subset of any language, this domain of language use is one of the most diverse, as it encompasses an incredibly wide array of text types, genres, topics, styles and levels of formality. Daily newspapers, in particular, with their wide circulation and readership, are often assumed to use the language "characteristic of the respective period and society they are published in" (Rademann, 2008: 49). As such, newspaper English of a particular country or region can offer valuable insights into the variety, as it is used by the community in general.

One of the ways in which the MEN Corpus has been exploited is in corpus-based lexicography of ME. The distinctiveness of the vocabulary of ME has been described and theorised in numerous studies (e.g., Tongue, 1974; Platt and Weber, 1980; Lee 1998, and so on), but until Higgleton and Ooi's (1997) *Times-Chambers Essential English Dictionary*, there had not been any attempt to codify the lexical features of ME in a dictionary. Ooi (2001) proposed several reasons for this. Chiefly among them are "local educational and political concerns that the wholesale adoption of the endonormative variety through its codification in the dictionary will inevitably lead to a decline in language standards" (Ooi, 2001: 168). In this paper, I argue that this does not have to be the case, and that in fact the codification of the lexical features of ME can be an effective way to signal the legitimacy of these features and the coming of age of this variety of English.

Although ME is internally varied (see, for instance, Baskaran's (1994) description of the three sociolects of ME), much of the research available today focuses on linguistic features that are the most distinctive—those that diverge most radically from the so-called "native" varieties of English (Newbrook, 1997: 229). These features are conspicuous, but they are often confined to colloquial speech or the basilectal variety of ME. Consequently, we do not have much information about the educated or formal variety. I argue that the use of a newspaper corpus offers us the opportunity to address this gap in our knowledge. Like newspaper Englishes all over the world, newspaper English in Malaysia tends to converge towards an international standard, yet conveys a distinct linguistic identity (Crystal, 1994: 24). The presence of a large amount of spoken data in the form of quotations (even if edited) provides another dimension to this relatively formal domain of language use. Malaysian English newspapers are a rich source of institutionalised lexical changes. These features are not ephemeral innovations, but have become entrenched in the linguistic system of ME as a result of widespread use. They are produced and understood by the community,

and are generally considered acceptable for use across a wide range of intranational domains. These characteristics of Malaysian English newspapers make them an invaluable source of data for the creation of dictionary entries that are relevant to the educational and communicational needs of the society while maintaining an endonormative standard. This article describes the analysis of the MEN Corpus for lexicographic evidence and the creation of dictionary entries based on data extracted from the corpus and from other secondary sources, focusing on the influences of Malay and Chinese languages on ME.

MALAYSIAN ENGLISH NEWSPAPER CORPUS

The MEN Corpus is a five-million-word corpus of newspaper articles published between 1 August 2001 and 30 January 2002. The sources of data are *The Star* and the *New Straits Times*, two of the most authoritative and widely read English language dailies in Malaysia. These newspapers are also the oldest surviving English-language newspapers in the country: The *New Straits Times* was established on 15 July 1845, while *The Star* was first published on 9 September 1971. At the time when the MEN Corpus was compiled, these two publications had a larger breadth of articles as well as a wider scope of coverage compared to other English-language newspapers in Malaysia.

In total, 91 issues of *The Star* and 61 issues of the *New Straits Times* were selected for sampling. To ensure that the corpus was representative of Malaysian newspaper English, several text categories were excluded. Among these were news, sports and business reports purchased from international newswire agencies (e.g., *Associated Press*, *Agence France-Presse* and *Reuters*), reprints of features from foreign sources, special-interest pieces contributed by foreign writers, and letters and opinion pieces written by individuals with names not immediately recognisable as Malaysian. In addition to this, illustrations, photographs, television guides, lottery results, poetry, cartoons, advertisements and classifieds were also excluded to ensure that features chosen for study could be examined in a fair amount of meaningful context. The final version of the corpus comprises a balanced spread of text categories which includes local and national news stories, regional and international news reportage by Malaysian correspondents overseas, news stories produced by *Bernama*—the Malaysian National News Agency, court reports, parliamentary reports, opinion pieces (including comments, letters to the editor and editorials), business and financial news, sports news, and features.

ANALYSING THE MALAYSIAN ENGLISH NEWSPAPER CORPUS FOR LEXICOGRAPHIC EVIDENCE

WordSmith Tools

The MEN Corpus has been successfully analysed using several corpus analysis software packages, including TACT, MonoConc and WordSmith Tools. However, as the vast majority of the analyses reported here were performed using WordSmith Tools, I shall begin by describing how this software package was used to extract the data required.

Developed by Mike Scott, WordSmith Tools is a suite of three programmes—WordList, KeyWords and Concord. Each of these serves specific functions that assisted me in identifying and extracting the raw data needed for interpretations of word meaning and use. Using WordList, I was able to generate two lists of all the types that occur in the MEN Corpus—an alphabetically-ordered one and a frequency-ordered one. Preliminary surveys of these word lists gave me an idea of the lexical features worthy of further investigation. KeyWords allowed me to identify words that are unusually frequent in the MEN Corpus. The programme does this by comparing the MEN Corpus word list with a larger reference word list. For the present purpose, I used the ten-million-word newspaper component of the British National Corpus as a reference corpus. The lists generated by these two programmes were useful for the identification of distinctive lexical features, in particular those that bear the influences of local languages overtly. Malay and Chinese loanwords (e.g., *bunga manggar*, *nasi lemak*, *amah* and *ang pow*) and compound blends (e.g., *pandan mat*, *tidak apa attitude*, *Chinese sinseh* and *popiah skin*) as well as their derived and inflected forms (e.g., *datukship* and *kiasu-ism*) are especially visible.

Not all lexical features can be identified through WordList and KeyWords. In addition to loanwords and compound blends, the vocabulary of ME has also been enriched by the presence of loan translations and lexical creations. The former is a category of lexical features that result from the literal translation of local compounds (e.g., *rice bowl* and *red bean paste*) while the latter comprises innovative words and phrases that have no known model in any of the local languages. While loan translations are purely English in form, lexical creations may be English (e.g., *dry kitchen* and *five-foot way*) or hybrid (e.g., *balik kampung* rush and *kopitiam* table) in form. Features that are entirely English in form are not immediately identifiable from a word list or a keyword list. For these features, I had to rely on earlier studies and on my own intuition as a native speaker of ME to track them down. Following Sinclair (1991: 18), I selected high frequency words and phrases over low frequency ones, and avoided *hapax legomena* or items which occur only once.

Once these lexical features were identified, they had to be analysed in context. This required the use of Concord. Concord is a tool that looks for all

instances of a particular search word in a corpus and presents each of them in the form of a concordance line which comprises the search word and a pre-specified amount of context. The entire set of concordance lines is presented in a concordance display. A particularly convenient feature of Concord is its ability to perform wildcard searches. By incorporating symbols like an asterisk, a question mark or a slash in a search word, it is possible to simultaneously search for a whole range of words that have certain common characteristics. For example, a search for one of the keywords of the MEN Corpus *datuk** yielded not only all instances of *datuk*, but also ten instances of the plural form of the word *datuks*, and 4 instances of the derived form *datukship*. Based on the concordance lines of a particular search word, the following information was deduced: (1) the meaning(s) of the word as used in ME; (2) the word class(es) assigned to it; (3) derived and inflected forms of the word and their meanings; and (4) citations which illustrate its different usages.

Dictionaries as Secondary Sources

In addition to the information derivable from context, the creation of ME dictionary entries required etymological information. A total of twelve old and contemporary dictionaries and glossaries were used for this purpose:

1. *A Grammar and Dictionary of the Malay Language* (Crawfurd, 1852);
2. *A Malay-English Dictionary* (Wilkinson, 1959);
3. *Dwibahasa Kamus Delta* [Delta Bilingual Dictionary] (Lufti and Awang, 1993);
4. *Kamus Melayu Global* [Malay Global Dictionary] (Hasan, 1997);
5. *Kamus Dewan Edisi Ketiga* [Third Edition of the Dewan Dictionary] (Noresah et al., 2000);
6. *Kamus Lanjutan Bahasa Malaysia-Bahasa Inggeris* [Advanced Malay-English Dictionary] (Abd. Aziz, 2003);
7. *Loan-words in Indonesian and Malay* (Jones, 2007);
8. *The Cantonese Speaker's Dictionary* (Cowles, 1965);
9. *Chinese-English Dictionary* (Mathews, 1972);
10. *The Chinese-English Dictionary* (English Department of the Beijing Foreign Languages Institute, 1981);
11. *Putonghua-Southern Fujian Dialect Dictionary* (Chinese Dialect Research Group under the Chinese Language and Literature Research Institute of Xiamen University, 1982); and
12. *Hobson-Jobson* (Yule and Burnell, 1903).

These secondary sources allowed for a more comprehensive analysis of the etymology of the lexical features chosen for codification.

Creating a Dictionary Entry

In order to illustrate how the information deduced from the concordance lines and the dictionaries was brought together in a dictionary entry, I shall describe the analysis of the lexical feature *kongsi*. A search for *kongsi** in the MEN Corpus yielded 22 instances of *kongsi* and 2 instances of the plural form *kongsis* (see

Figure 1). The word *kongsi* has its origin in Chinese 公司 (Hokkien *gong si*, Cantonese *kung sz* and Mandarin *gong si*—literally "joint control"). Originally, this word was used to refer to a collective labour group or joint ownership of a cargo. Subsequently, it also came to be used in reference to surname associations. In contemporary Chinese, the most common meaning for this word is "public company" (Mathews, 1972: 542). The word was borrowed into Malay, most likely during the 19th-century expansion of the tin-mining industry in Malaya, and in his widely-acclaimed *Malay-English Dictionary*, Wilkinson defined the word as follows:

kongsi. Ch. Partnership or association of any sort. Usually a Chinese guild or secret society, but also of syndicates in general (**përkongsian**, Pert. Tebu 9) and even of a British missionary body (Ht. Abd. 118).
Képalá k.: secret-society headman. **Rumah k.:** house for gang of Chinese coolies (Wilkinson, 1959: 610).

Based on the concordance lines of *kongsi* and information obtained from various dictionaries, it was deduced that in ME: (1) *Kongsi* has two main meanings—"corporate body, often based on surname affiliation" (lines 5–12, 14–17, 20, 21) and "shared accommodation for (migrant) labourers usually found at the worksite" (lines 1–4, 13, 18, 19, 22–24); (2) *Kongsi* is not generally used to denote the contemporary meaning of Chinese 公司 "public company"; (3) *Kongsi* is used as a count noun, and it can be inflected with the morpheme *-s* to indicate plurality; and (4) *Kongsi* is likely borrowed from Hokkien as the transliteration replicates the Hokkien pronunciation of 公司.¹ Using this information, the dictionary entry below was created for the Chinese loanword *kongsi*:

kongsi *n.* Pl. *-s*. [Hokkien 公司, lit. "joint control"] **1** Corporate body, often based on surname affiliation. **2** Shared accommodation for (migrant) labourers usually found at the worksite.

1 2001 *The Star* 29 Nov. PENANG: It was like a reunion of long lost brothers for the Toh clansmen of Penang and Johor who met for the first time here on Tuesday. Ironically, the 109-year-old Penang Toh Kongsi and the 24-year-old Toh Association Malaysia ... were unaware of each other's existence until last year. **2 2002** *The Star* 12 Jan. ... the girl's father lodged a police report against the immigrant worker, who was arrested at a kongsi in the orchard.

Using this method, 457 dictionary entries were created (for the entire glossary, see Tan, 2013: 155– 206). Consisting primarily of nouns, these features include food names, terms for festivals and religious practices, names of social and recreational activities, and honorific titles and terms of respect.

1	of another accomplice – a 32-year-old Indonesian – at a		
2	Mon. - Two gangs clashed with each other at a		
3	against the immigrant worker, who was arrested at a		
4	Ampong, last Wednesday. Several youths went to a		
5	ties among them. "We have been quite an introvert		
6	toured the trail from the Syed Alatas Mansion to Khoo		
7	by restoring its old premises in Beach Street. The Khoo		
8	rented out to non-clansmen. In recent years, the Khoo		
9	KHOQ		
10	THE Khoo		
11	will reopen its doors to tourists soon. A visit to the Khoo		
12	a visit to the newly-restored temple yesterday, Khoo		
13	if illegals are detained, we cannot take them back to their		
14	here on Tuesday. Ironically, the 109-year-old Penang Toh		
15	were greeted by their Penang brothers outside Wisma Toh		
16	and hope to build a greater rapport in future." Penang Toh		
17	long lost brothers", said: "We discovered the Penang Toh		
18	selected areas such as squatter settlements and workers'		
19	illegal immigrants, including 17 women, were arrested in		
20	houses. An exciting new chapter is unfolding for the Khoo		
21	Penang Toh Kongsî secretary Charles Toh said the		
22	as the founder of the Ariffin Mosque and Mesjid		
23	of such workers in Putrajaya and they daringly stay in		
24	from the place being wounded. The bazaar serves several		
		kongsî	in Damansara Utama," he added. Police have already
		kongsî	in Tasik Permai, Ampang, last Wednesday. Several youths
		kongsî	in the orchard. According to the report, the rape was
		kongsî	occupied by Myanmar nationals and demanded RM5 from
		kongsî	that has kept a low profile. "We only learnt of the
		kongsî	and the Acheen Street Mosque. From the mosque, they
		kongsî	and the Kapitan Keling Mosque are two more examples of
		kongsî	has embarked on a revitalisation process, winning national
		kongsî	in Cannon Square, George Town, Penang, is unique not
		kongsî	Temple at Cannon Square in Penang is looking even more
		kongsî	temple will provide tourists a detailed insights into the
		kongsî	Trustee Khoo Kay Hock said the workers were assigned to
		kongsî	to collect their things," he said. According to the
		kongsî	and the 24-year-old Toh Association Malaysia (TAM)
		kongsî	at C'intra Street and later adjourned for a reunion dinner at
		Kongsî	secretary Charles Toh said the kongsî, also known as
		Kongsî	three years ago but we had no opportunity to meet until
		kongsî	as well as conducting roadblocks. In reply to Senator
		kongsî,	shophouses, kampungs and rented rooms in Plentong,
		Kongsî,	but the new development is bringing to an end the
		kongsî,	also known as Cheng Luan Seah (Toh Si), only took in
		Kongsî.	He is said to have arrived in Penang with Captain Francis
		kongsî	and the surrounding jungles. Somebody should take a
		kongsî	at worksites in Dengkil, Putrajaya and Cyberjaya. They

Figure 1: Concordance lines of "kongsî" from the MEN Corpus

Etymology

The 13 dictionaries were indispensable in the compilation of etymological information for the dictionary entries. This was particularly true for Malay loanwords. In modern Peninsular Malaysia, there are at least four major Malay dialects—the northwestern dialect spoken in Kedah, Perlis and Penang; the northeastern dialect spoken in Kelantan; the eastern dialect spoken in Terengganu; and the southern dialect spoken in Johor, Melaka, Pahang, Selangor and Perak (Asmah, 1977). Other minor Malay dialects used in Malaysia include those spoken in Sabah and Sarawak, some which originated in parts of present-day Indonesia, Creole Malay spoken by the Straits Chinese community in Melaka known locally as Baba Malay, and a bazaar variety—*bahasa pasar* "market language"—that is widely used as an inter-ethnic vernacular especially in commercial environments. Generally dialectal variation in Malay is not reflected in the lexicon of ME, but there are a number of loanwords which are distinctly derived from Javanese (e.g., *gamelan*, *soto* and *wayang*), Kedah Malay (e.g., *songket*) or Kelantan Malay (e.g., *budu*). In such cases, I noted the original source language of the feature in the entries, as shown below:

budu *n.* [Malay, orig. Kelantan Malay] A condiment made by pickling anchovies in brine; popular in the east coast states of Peninsular Malaysia.

2001 *New Straits Times* 22 Nov. Among the fare served for the hotel's Ramadan buffet dinner are various types of kerabu, ulam-ulaman, sambal belacan, budu and tempoyak.

songket *n.* Also **kain songket**. [Malay, orig. Kedah Malay] A traditional Malay hand-woven fabric with gold and silver threads, usually worn during official functions and ceremonies.

2001 *New Straits Times* 5 Sept. Some of the great crafts produced were the award-winning kain songket which has broken away from the usual repetitive motif and instead has a flowing floral motif that runs boldly across the material...

wayang *n.* [Malay, orig. Javanese] A local theatrical performance.

Comb.: **wayang kulit** shadow play, theatrical performance where shadow images are projected before a backlit screen; **wayang peranakan** theatrical performance where the characters speak the Peranakan language.

2001 *The Star* 2 Oct. A cultural theatre showcasing Chinese operas, wayang kulit and traditional music is more appropriate and so much more tasteful.

Besides dialectal variation, the Malay language also exhibits the influences of other languages with which it came into contact during its evolution in Archipelagic Southeast Asia. In the "dissertation" of his two-volume work, *A Grammar and Dictionary of the Malay Language*, Crawford noted the following:

An examination of the 4074 radical words of the Dictionary shows that the Malay language is composed of the following lingual elements: Native Malay words, 2003; common to the Malay and Javanese, 1040; Sanskrit, 199; Tâlugu or Telinga, 23; Arabic, 160; Persian, 30; and Portuguese, 19; which, in a 1000 words, give the following proportions respectively: Native, 491; Javanese, 255; Sanskrit, 49; Tâlugu about 5½; Persian, about 7; and Portuguese, about 4½ (Crawfurd, 1852: xiii).

The influences of these languages on Malay are certainly visible in the vocabulary of ME. For instance, Malay loanwords to do with the government, administration and the monarchy (e.g., *Menteri Besar*, *Raja*, *Raja Permaisuri Agong* and *Yang di-Pertua Neg(e)ri*) often incorporate Sanskrit morphemes. Even more conspicuous are the borrowing into ME through Malay of Arabic words related to Islam, such as *halal*, *haram*, *imam*, *kadi*, *khalwat*, *madrrasah*, *ulama* and *wali*. More recently, influences of southern Chinese (e.g., *kuih*, *mee goreng*, *nyonya* and *teh tarik*) and vernacular Indian languages (e.g., *mamak*, *nasi briyani*, *putu beras*, *roti* and *sambal*) on Malay have also played a role in shaping the lexicon of ME. As the impact of these languages, especially Sanskrit and Arabic, on ME is only indirect, with Malay being the apparent conduit through which their influence has been manifested, dictionary entries for these features indicate "Malay" as the source language, but include a note on the original source language. The entries below are some examples:

kadi *n.* [Malay, orig. Arabic *qadi*] A judge in Islamic affairs.

2001 *New Straits Times* 31 Oct. Under the ruling, couples wanting to get married have to undergo the HIV test, the results of which must be handed to the kadi before the wedding.

kuih *n.* Pl. same, **kuih-muih**, **kuih-kuih**. [Malay, orig. Hokkien *kóe*] Any type of local cakes, puddings, biscuits, pastries and fritters, made variously from glutinous rice flour, rice flour, wheat flour, cane sugar, palm sugar, coconut milk, grated coconut and eggs.

2001 *New Straits Times* 6 Nov. For dessert, check out ais kacang, assorted ice cream, bread and butter pudding, iced longan and jelly, kuih, French pastries and fresh fruits.

Comb.: **kuih bahulu** small cupcake, traditionally baked over charcoal fire in cast iron moulds; **kuih bangkit** light biscuit made of coconut milk, sugar and rice flour; **kuih kapit** thin wafer made from wheat flour, coconut milk, sugar and eggs.

Menteri Besar *n.* Pl. **Menteris Besar**. [Malay, orig. Sanskrit *mantri* "official" + Malay *besar* "big"] Chief Minister for any of the nine former Federated and non-Federated Malay states.

2001 *The Star* 19 Aug. The Menteri Besar has been so pleased with Yee's book that the state government purchased over 600 copies to be distributed to students in Sabak Bernam.

sambal *n.* [Malay, orig. Tamil *sambaar*] Spicy condiment made variously from chillies, tamarind, shrimp paste, etc.

2001 *New Straits Times* 28 Nov. "Over 70 per cent of our stay-in guests are local business travellers, thus even our food and beverage outlets cater for them. The coffee house serves mainly Malay food because locals tend to miss their sambal and 'warong' dishes."

Comb.: **sambal belacan** condiment of pounded chillies and dried shrimp paste; **sambal ikan bilis** condiment of chilli paste and dried anchovies; **sambal petai** condiment of chilli paste and petai.

The vast majority of ME features of Chinese origin have been borrowed from Hokkien, Cantonese and Mandarin. Although these are sometimes regarded as dialects of the same language, in actual fact there are significant differences in their vocabularies, pronunciation and syntax. These differences are often significant enough to allow us to identify the precise Chinese language from which an ME feature has been borrowed. We know, for instance, that the loanword *dim sum* "traditional Chinese snacks of dumplings, buns and sweets" is borrowed from Cantonese 點心 because the transliteration *dim sum* replicates the Cantonese pronunciation of the term. In Mandarin, the transliteration is *dian xin* while in Hokkien it is *diam sim*. This study foregrounds these differences by providing, where possible, in-depth etymological information—the specific source language, original Chinese characters and their literal meaning—for the features identified. The following are some examples:

samfoo *n.* [Cantonese 衫褲, lit. "dress trousers"] A suit consisting of a blouse and a pair of loose trousers, traditionally worn by Chinese women.

2001 *The Star* 8 Oct. Kicking off with two Chinese women dressed in Indian garb followed by a couple of Indian women dressed in samfoo, it showcased a collaboration which represented the various races in a truly unified manner while epitomising the rich diversity which was unique to Malaysia.

towkay *n.* Pl. -s. [Hokkien 头家, lit. "head family"] **1** Chinese businessman or shop owner. **2** A term used to refer to the oldest or most experienced member of a particular group.

1 **2001** *New Straits Times* 6 Nov. ...Andy said: "I wanted to know if the towkay (owner) had repaired the door and if he did, I wanted to ask him if he had seen Chindra or not." **2** **2001** *The Star* 5 Aug. ...the former stars of Malaysian and Singapore football showed they can still dance on the big stage. The Selangor veterans had household names like Datuk M. Chandran, Abdul Rashid Hassan, Ismail Ibrahim, K. Gunalan, "Towkay" Datuk Soh Chin Aun, ...

wushu *n.* [Mandarin 武术, lit. "martial skills"] A form of Chinese martial arts.

2002 *The Star* 20 Jan. Wushu promotes physical well-being and a healthy body promotes a healthy mind which in turn builds personality and good behaviour.

The borrowing of Chinese words into written ME involves representing the pronunciation of the imported morphemes using Latin script. Since there is no standard system of transliterating morphemes of Chinese origin in ME, we see a

fair amount of variability in the orthographic representations of these words. For example, *koay teow*, *kuay teow*, *kuey teow* and *kway teow* are all representations of the Hokkien pronunciation of 粿条 "rice noodles"; while *wantan*, *wanton* and *wonton* are representations of the Cantonese pronunciation of 馄饨 or 云饨 "dumpling stuffed with minced meat or prawns." The dictionary entries order these variants according to usage preference:

wantan *n.* Also **wanton**, **wonton**. [Cantonese 馄饨 or 云饨, lit. "stuffed dumpling"]
Dumpling stuffed with minced meat or prawns.

wantan_noodles [Cantonese 云饨面, lit. "stuffed dumpling noodles"]
Noodles served with wantan.

2001 *New Straits Times* 26 Aug. Fresh wantan noodles, these are reminiscent of those I had in Hongkong with that al dente bite and not in the least bit overdone.

Semantic Adaptations

An important facet of lexical borrowing in ME is the semantic adaptation of loanwords, particularly Malay loanwords. Malay words do not always retain their original meanings in the process of being imported into ME. Often only a single sense is transferred, and sometimes the meaning in ME conveys a cultural specificity that is absent in the original range of meanings. Table 1 illustrates this by providing a comparison of the meanings of several words in Malay and in ME, as represented by the MEN Corpus. In the borrowing of the words *dadah*, *gatal*, *kacang*, *roti* and *ulu*, semantic restriction occurs. For example, the word *gatal* in Malay can mean both "itchy" (a sensation) and "mischievous and flirtatious" (commonly associated with lecherous men), but in ME it appears to be used only to express the latter meaning. Similarly, the Malay word *kacang* has several senses—"peas," "beans," "lentils" and "nuts"—but in ME the word most often refers to "roasted nuts, usually eaten as a snack." In the borrowing of these words, only one specific sense seems to have been transferred, usually a sense that cannot be concisely expressed using existing English words.

Table 1: Semantic adaptation of Malay loanwords in Malaysian English

Lexical item	Meaning(s) in Malay	Meaning(s) in ME
<i>dadah</i>	<i>n.</i> 1 A substance used as a component of a medication. 2 A chemical substance, such as a narcotic or hallucinogen, that affects the central nervous system.	<i>n.</i> A chemical substance, such as a narcotic or hallucinogen, that affects the central nervous system.
<i>gatal</i>	<i>a.</i> 1 Itchy. 2 Mischievous and flirtatious, usually of lecherous men.	<i>a.</i> Mischievous and flirtatious, usually of lecherous men.
<i>kacang</i>	<i>n.</i> Peas, beans, lentils or nuts.	<i>n.</i> Roasted nuts, usually eaten as a snack.
<i>rakyat</i>	<i>n.</i> 1 The citizens of a state or country. 2 The common people (as opposed to the government or the aristocracy).	<i>n.</i> The common people (as opposed to the government or the aristocracy) of Malaysia.
<i>rotan</i>	<i>n.</i> 1 Any of various climbing plants of tropical Asia, having long, tough, slender stems. 2 The stems of any of these plants, used to make wickerwork, canes, and furniture. 3 A cane made from these plants. 4 Judicially-sanctioned caning in Malaysia.	<i>n.</i> 1 A rattan cane used for inflicting judicially-sanctioned corporal punishment in Malaysia. 2 Any of various climbing plants of tropical Asia, having long, tough, slender stems. 3 (<i>rare</i>) A cane. 4 (<i>rare</i>) Judicially-sanctioned caning in Malaysia.
<i>roti</i>	<i>n.</i> Bread	<i>n.</i> Bread, usually the local version of a white loaf, which is slightly sweet and has a very soft texture.
<i>ulu</i> <Malay <i>hulu</i> >	<i>n.</i> 1 The source of a river. 2 Inland area. 3 The handle of a tool, knife, etc.	<i>n.</i> A provincial place, back country.

In some cases, the transfer from Malay to ME gives the loanwords a new cultural specificity that is absent in their original meanings. I shall illustrate this point using the loanword *rotan*. Figure 2 is the edited² concordance display of *rotan* from the MEN Corpus. There are four main senses of *rotan* in Malay (see *rotan* in Table 1) and three of these appear in the MEN Corpus. The reference to "any of various climbing plants of tropical Asia, having long, tough, slender stems" is seen in lines 2 and 3 of the concordance display, the reference to "a cane" in line 1, and the reference to "judicially-sanctioned caning" in line 12. The primary sense of *rotan* in ME, however, is "a rattan cane used for inflicting judicially-sanctioned corporal punishment in Malaysia," and this sense occurs 19 times in the MEN Corpus (see lines 4–11 and lines 13–23 in Figure 2). This sense of *rotan* in ME has a cultural specificity—not just any cane but one that is used to

carry out court-ordered canings—and it is this specificity that is not evident in the range of meanings of *rotan* in Malay. The reference to this meaning is encapsulated in the phrase *stroke(s) of the rotan*, which is a localised version of the English phrase *stroke(s) of the cane*—the English phrase is retained but for the substitution of *rotan* for *cane*.

1	-year-old Teoh Lee Sean by using a	rotan	between Oct 2000 and July 10 last
2	rotan manau (Calamus manan) and	rotan	sega (Calamus caesius) grew better
3	in Negri Sembilan found that	rotan	manau (Calamus manan) and rotan
4	12 months and given one stroke of the	rotan	for possessing a parang. On a third
5	to 15 years' jail and 10 strokes of the	rotan	for dadah possession after hearing
6	ordered to be given 10 strokes of the	rotan	for outraging the modesty of a
7	to 15 years' jail and 10 strokes of the	rotan	for dadah possession. The court
8	today imposed five strokes of the	rotan	on a 27-year-old odd-job worker who
9	nine years' jail and nine strokes of the	rotan	after he pleaded guilty to nine
10	to 12 years' jail and six strokes of the	rotan	for rape and two months' jail for
11	years' jail and three strokes of the	rotan	after he pleaded guilty to sodomising
12	to a maximum 20 years jail and	rotan	for the offence. Investigating officer
13	or jailed and given six strokes of the	rotan,"	he said when debating amendments
14	five years' jail and three strokes of the	rotan,	the court found it difficult to decide
15	years and no less than six strokes of	rotan.	Mohd Salleh and Sait also face
16	with three years' jail and strokes of the	rotan.	Meanwhile, offences under Firearms
17	14 years' jail and six strokes of the	rotan.	Wan Afrah fixed three days
18	14 years' jail and six strokes of the	rotan.	Zainuri, who was believed to have
19	and no less than six strokes of the	rotan.	Sait is facing an alternative charge of
20	to six years' jail and six strokes of the	rotan.	Jujili @ Samrin Gali pleaded guilty
21	years and receiving six strokes of the	rotan.	"We are looking for several
22	jail and a minimum six strokes of the	rotan.	Both also face a second charge of
23	jail and no less than six strokes of the	rotan.	

Figure 2: Concordance lines of "*rotan*" from the MEN Corpus

The same cultural specificity is observed in the meaning of *rakyat* in ME. In Malay, the word *rakyat* has two main senses: "the citizens of a state or country" and "the commoners (as opposed to the government or the aristocracy)." If we examine the concordance lines of *rakyat* from the MEN Corpus (see Figure 3), we see that in all 45 lines, *rakyat* refers to "the common people (as opposed to the government or the aristocracy) of Malaysia." Hence, the referent for *rakyat* is more specific in ME than in the Malay language.

1	with the welfare of the staff and	rakyat	in mind. Q: Your name came out
2	together afterwards. The	rakyat	cannot be influenced into not
3	creates an impression among the	rakyat	on who is right and who is
4	simply to serve the party and the	rakyat	to the best of our abilities," he
5	help foster better ties between the	rakyat	and their elected representatives.
6	This must be understood by the	rakyat	and we are aware of the
7	"What is being scrutinised by the	rakyat	is the quality of service concerni
8	MPs, the people elected by the	rakyat	to give voice to their aspirations
9	"What is being scrutinised by the	rakyat	is the quality of service
10	in their constituencies. The	rakyat	can also reach State Executive
11	ds: "This is an opportunity for the	rakyat	to express their gratitude to
12	top shopping destination, it is the	rakyat	(people) who will determine the
13	the <i>jihad</i> required was making the	rakyat	aware that only Islam could
14	time back, some segments of the	rakyat	were prepared to ask the US to
15	the country and the tenacity of the	rakyat	to rise above the trials of life,
16	sonal interest in the welfare of the	rakyat	saw him venturing into the
17	derstanding and co-operation, the	rakyat	must have faith and confidence
18	doing well. If the consumer or the	rakyat	at large have enough confidence
19	elected by the people to serve the	rakyat	and the Government. A
20	overnment does have to teach the	rakyat	that smoking is supposedly bad
21	combat it. At the same time, the	rakyat	also ask that the US provides
22	scary. We need to explain to the	rakyat	what it really is. What we want
23	afforded them by turning the	rakyat	against the Government. This
24	the back." Later, speaking at the	meet-the-rakyat	session at Stadium Perpaduan
25	Wan Ismail: The Budget is a	rakyat's	budget. PAS president Datuk
26	opposed the move as the	rakyat's	money shouldn't be risked in
27	nothing of throwing sand in the	rakyat's	rice bowl. The second is
28	to be more sensitive to the	rakyat's	problem when approving any
29	as a mother uniting her children	(rakyat)	who each has their own
30	for real. It's quite sad. As the	rakyat,	all we can do now is offer our
31	they (PAS) going to convince the	rakyat,	especially us (UMNO) on the
32	out" to improve efficiency for the	rakyat,	implementation by local
33	to duty and his empathy for the	rakyat,	some of which gained
34	people, especially the Selangor	rakyat.	He always reminded us how
35	vernmental organisations and the	rakyat.	Dennis J. Parker of Middlesex
36	e that was not going to benefit the	rakyat.	"If they rule this country, the
37	should be utilised to benefit the	rakyat.	When undertaking mega or big
38	was generally lauded by the	rakyat.	Dr Mohd Khir's latest decision
39	d been felt most keenly by the	rakyat.	"It's hoped that with new
40	educational opportunities for the	rakyat.	Doubtless, this package helps
41	at look up to the aspirations of the	rakyat.	And also to promote forward
42	way against the wishes of the	rakyat.	"The 19 had gone to the extend
43	the MPs were representing the	rakyat.	"If we Yang Berhormat find it
44	he was always close to the	rakyat.	When he was appointed King,
45	arty will explain its concept to the	rakyat.	"Often, the image is one that is

Figure 3: Concordance lines of "rakyat" from the MEN Corpus

In the absence of the complete range of meanings that these words have in their original language, it is assumed that semantic restriction has occurred as these words are borrowed into ME. As can be seen below, the entries for these loanwords include only the meanings found in the MEN Corpus:

rakyat *n.* [Malay, orig. Arabic *ra'iyya* "subjects"] The common people (as opposed to the government or the aristocracy) of Malaysia.

2001 *New Straits Times* 28 Aug. For Ayub, this statement reflects the capricious ideologies spread by certain quarters who take advantage of the leeway afforded them by turning the rakyat against the Government.

rotan *n.* [Malay] *n.* **1** A rattan cane used for inflicting judicially-sanctioned corporal punishment in Malaysia. **2** Any of various climbing plants of tropical Asia, having long, tough, slender stems. **3** (*rare*) A cane. **4** (*rare*) Judicially-sanctioned caning in Malaysia.

1 2001 *New Straits Times* 5 Sept. In allowing Tan's appeal, the court sentenced him to 15 years' jail and 10 strokes of the rotan for dadah possession.

3 2002 *The Star* 24 Jan. ... are jointly charged with voluntarily causing grievous hurt to nine-year-old Teoh Lee Sean by using a rotan between Oct 2000 and July 10 last year at Jalan Tiram in Cheras. **4 2001** *New Straits Times* 18 Aug. He also said he understood that he could be sentenced to a maximum 20 years jail and rotan for the offence.

Morphosyntactic Adaptations

In addition to semantic adaptations, loanwords are also subject to morphosyntactic adaptations. ME users sometimes create new grammatical categories and novel word forms by adding English inflectional and derivational morphemes to words of Malay and Chinese origin. By far the most productive inflectional morpheme in ME is the plural *-s* affix, which is regularly utilised to indicate plurality in count nouns. Examples of loanwords that have been given a plural form are *ang pow* (plural *ang pows*), *bomoh* (plural *bomohs*), *cheongsam* (plural *cheongsams*) and *pondok* (plural *pondoks*). ME users also often employ derivational morphemes to create new words. The most common of these are: (1) the prefix *non-* indicating "not" (e.g., *non-halal* "not *hahal*"); (2) the suffixes *-ean*, *-an*, *-ese*, *-ite* and *-ian* indicating "residents of a state" (e.g., *Johorean* "resident of *Johor/Johore*," *Kedah* "resident of *Kedah*," *Kelantanese* "resident of *Kelantan*," *Penangite* "resident of *Penang*," and *Sarawakian* "resident of *Sarawak*"); (3) the suffix *-ship* indicating "a state or condition" (e.g., *Datukship* "the state of having been awarded the title of *Datuk*"); and (4) the suffix *-ism* meaning "a distinctive practice associated with a particular group" (e.g., *kiasuism* "the actions or conduct of people who are *kiasu*"). The entries below show how some of these morphosyntactic adaptations have been codified:

- cheongsam** *n.* Pl. -s. [Cantonese 长衫, lit. "long dress"] A form-fitting dress characterised by a Mandarin collar and Chinese-craft buttons, popular among Chinese women.
2002 *The Star* 30 Jan. The fashion range at Parkson, which is priced from RM43.90, varies from relaxing Capri pants to blouson blouses to the more elegant dresses and cheongsams.
- halal** *a.* [Malay, orig. Arabic *ḥalāl* "permitted"] Permitted under Islamic law, usually in reference to food.
2001 *New Straits Times* 4 Nov. Of the switch to halal Chinese cuisine, he says that he and his kitchen team have been preparing for it for months, testing various ways to get similar or almost similar flavours and textures.
non-halal not permitted under Islamic law, usually in reference to food and restaurants.
2001 *The Star* 15 Aug. Shops that do not have a Chinese restaurant permit, even though they are Chinese-owned and have been selling other non-halal food like pork, cannot legally sell beer without the permit.
- Kelantan** *n.* [Malay] A state situated in the northeast region of Peninsular Malaysia.
Kelantanese *n.* Pl. same. A native or inhabitant of the state of Kelantan.
2001 *New Straits Times* 19 Sept. Kelantanese are renowned for their warmth and friendliness.
- kiasu** *a.* [Hokkien 惊输, lit. "afraid of losing"] Being afraid of missing out on opportunities, losing out to other people, etc.
2001 *The Star* 16 Sept. ...Paul Jambunathan said aggressive driving has become the nation's number one transportation problem and that Malaysian drivers have gained international notoriety for being selfish and kiasu, and for defying law and authority when on road.
kiasu-ism *n.* The actions or conduct of people who are kiasu.
2001 *The Star* 19 Aug. Our neighbour Singapore, is being plagued by kiasu-ism as their society is breaking up into individuals who are focused only on their own personal needs and wants.

These morphosyntactic adaptations serve the important function of integrating these lexical features into the linguistic system of ME. In other words, although of non-English in origins, they exhibit the morphosyntactic characteristics of other, more established English nouns and adjectives.

CONCLUSION AND FUTURE EXPLORATIONS

This article demonstrates the value of a small newspaper corpus as a source of lexicographic evidence for the compilation of dictionary entries of localised words in ME. Clearly, the loanwords, compound blends, loan translations and lexical creations examined through this project are crucial for the linguistic and sociocultural needs of the community. Not only do they occur in the relatively formal register of newspaper English, they also exist in stable grammatical contexts which provide us with information about their meanings and usage.

Many of these features undergo morphosyntactic adaptations that allow them to be more fully integrated into the linguistic system of ME. The codification of these lexical features is obviously necessary if they are to be given due recognition as legitimate items of the vocabulary of ME, and if the legitimacy of ME is to be acknowledged.

Obviously, the features codified through this project represent only a small subset of distinctive vocabulary items of ME. Besides Malay and Chinese, the influences of languages such as Tamil, Iban and Kadazan will have to be examined. Beyond loanwords, compound blends, loan translations and lexical creations, the lexicon of ME has also been shaped by the semantic adaptations of many core English words. Some well-known examples of words that have been adapted include *auntie/aunty*, *uncle*, *follow*, *fetch*, *stay*, *take* and *wear* (for other examples, see Tan, 2013: 123–130). In ME, *auntie/aunty* and *uncle* are often used as terms of respect for older women and men, regardless of whether or not they are one's relatives. The verb *follow*, besides the inner-circle meaning of "to go or come after a person or other object in motion," is used in ME to mean "to accompany or to go with." Whether, and if so how, such lexical features can be codified is a question of future scholarship in this area should address. Beyond the fact that some of these meanings are clear products of over-generalisation associated with group second language learning, there is also the likelihood that some of these senses "may well be survivals of 19th-century British usages which have become obsolete or archaic in the mother-country itself" (Görlach, 1991: 27, cited in Ooi, 2001: 171). Future studies that examine whether such lexical features have stabilised in the context of ME and are therefore legitimate variants in an emerging endonormative standard will be necessary before we can even consider whether they should be included in a ME dictionary.

The fact that the MEN Corpus is a small corpus representing newspaper English has obvious implications on the features selected for codification, and the range of meanings found for each feature. It is however important to recognise that "any corpus, however big, can never be more than a minuscule sample of all the speech and writing produced or received by all the users of a major language on even a single day" (Kennedy, 1998: 66). The humble beginnings of the COBUILD English Language Dictionary—the COBUILD lexicographers started with a corpus of a corpus of 7.3 million words in the 1980s (Hank, 2012: 402)—is a useful reminder that even a small corpus can provide us the basic building blocks for a dictionary. The fact that today COBUILD dictionaries draw materials from the 2.5-billion-word Collins Corpus (of which the 650-million-word Bank of English is a part of) emphasises the importance of viewing any corpus-based lexicography project as an evolving, diachronic enterprise.

NOTES

1. Although 公司 is represented as *gong si* in both Hokkien and Mandarin in this article, the actual pronunciations of the word in the two languages are quite different. This difference is not reflected in the transliteration because this study utilises different romanisation systems for the two languages.
2. In several concordance lines, the word *rotan* appears as part of a proper noun. These lines have been deleted.

REFERENCES

- Abd. Aziz Rahman. 2003. *Kamus lanjutan Bahasa Malaysia-Bahasa Inggeris [Advanced Malay-English dictionary]*. Petaling Jaya: Federal Publications.
- Asmah Omar. 1977. *Kepelbagaian fonologi dialek Melayu [Diversity in the phonology of Malay dialects]*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Baskaran, L. 1994. The Malaysian English mosaic. *English Today* 10(1): 27–32.
- Chinese Dialect Research Group under the Chinese Language and Literature Research Institute of Xiamen University. 1982. *Putonghua-Southern Fujian dialect dictionary*. Fuzhou: Fujian People's Publishing House.
- Cowles, R. T. 1965. *The Cantonese speaker's dictionary*. Hong Kong: Hong Kong University Press.
- Crawford, J. 1852. *A grammar and dictionary of the Malay language; with a preliminary dissertation*. London: Smith, Elder and Co. http://openlibrary.org/books/OL24821279M/A_grammar_and_dictionary_of_the_Malay_language/ (accessed 15 October 2013).
- Crystal, D. 1994. What is standard English? *Concorde (English speaking union)*, 24–26. http://www.davidcrystal.com/DC_articles/English52.pdf/ (accessed 16 July 2012).
- English Department of the Beijing Foreign Languages Institute. 1981. *The Chinese-English dictionary*. Hong Kong: Commercial Press.
- Görlach, M. 1991. *Englishes: Studies in varieties of English, 1984–1988*. Amsterdam: John Benjamins.
- Hanks, P. 2012. The corpus revolution in lexicography. *International Journal of Lexicography* 25(4): 398–436.
- Hasan Hamzah. 1997. *Kamus Melayu global [Malay global dictionary]*. Shah Alam: Piramid Perdana.
- Higgleton, E. and V. B. Y. Ooi, eds. 1997. *Times-chambers essential English dictionary*. 2nd ed. Edinburgh and Singapore: Chambers Harrap and Federal.

- Jones, R., ed. 2007. *Loan-words in Indonesian and Malay*. Leiden, The Netherlands: KITLV Press.
- Kennedy, G. 1998. *An introduction to corpus linguistics*. London and New York: Longman.
- Lee, S. K. 1998. *Manglish: Malaysian English at its wackiest*. Kuala Lumpur and Singapore: Times Books International.
- Levin, M. 2006. Collective nouns and language change. *English Language and Linguistics* 10(2): 321–343.
- Lufti Abas and Awang Sariyan. 1993. *Dwibahasa kamus delta (Edisi KBSM) [Delta bilingual dictionary (KBSM edition)]*. Petaling Jaya: Pustaka Delta Pelajaran.
- Mathews, R. H. 1972. *Mathews' Chinese-English dictionary*. Cambridge, MA: Harvard University Press. (Reprinted from: 1931, Shanghai, China: China Inland Mission and Presbyterian Mission Press).
- Newbrook, M. 1997. Malaysian English: Status, norms, some grammatical and lexical features. In *Englishes around the World: Vol. 1: General studies, British Isles, North America – Studies in Honour of Manfred Görlach*, ed. W. S. Edgar, 229–256. Amsterdam/Philadelphia: John Benjamins.
- Noresah Baharom, Rusli Abdul Ghani, Mohd. Nor Abd. Ghani, Ibrahim Ahmad, Aziah Tajudin, Salmah Jabbar and Rahamah Othman, eds. 2000. *Kamus dewan edisi ketiga [Third edition of the dewan dictionary]*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Ooi, V. B. Y. 2001. Upholding standards or passively observing language?: Corpus evidence and the concentric circles model. In *Evolving identities: The English language in Singapore and Malaysia*, ed. V. B. Y. Ooi, 168–183. Singapore: Times Academic Press.
- Platt, J. and H. Weber. 1980. *English in Singapore and Malaysia: Status, features, functions*. Kuala Lumpur: Oxford University Press.
- Rademann, T. 2008. Using online electronic newspapers in modern English-language press corpora: Benefits and pitfalls. *ICAME Journal* 22: 49–72.
- Sinclair, J. 1991. *Concord, concordance, collocation*. Oxford: Oxford University Press.
- Tan, S. I. 2013. *Malaysian English: Language contact and change*. Frankfurt am Main: Peter Lang.
- Tongue, R. K. 1974. *The English of Singapore and Malaysia*. Singapore: Eastern Universities Press.
- Westin, I. 2002. *Language change in English newspaper editorials*. Amsterdam: Rodopi.
- Wilkinson, R. J. 1959. *A Malay-English dictionary (Romanised)*. London: Macmillan.
- Yule, H. and A. C. Burnell. 1903. *Hobson-Jobson: A glossary of colloquial Anglo-Indian words and phrases, and of kindred terms, etymological, historical, geographical and discursive*. London: John Murray.